THE MADS-BOX GENE FAMILY OF THE BASAL EUDICOT AND HYBRID AQUILEGIA COERULEA 'ORIGAMI' (RANUNCULACEAE)1

Bharti Sharma² and Elena M. Kramer²*

ABSTRACT

The MADS-box genes form a large family of pan-eukaryotic transcription factors that are involved in various aspects of plant growth and development, particularly reproduction. To understand the extent of their conservation and divergence in the emerging model genus Aquilegia L. (Ranunculaceae), we have annotated 47 MADS-box containing loci from the recently released hybrid A. coerulea E. James 'Origami' genome sequence. Phylogenetic analysis of these sequences along with those previously identified from Arabidopsis (DC.) Heynh. and Oryza L. demonstrates that we were able to recover members of all major subfamilies with the exception of clear MB representatives. The evolution of the Aquilegia type I loci is similar to what has been observed for other angiosperms in exhibiting relatively recent gene radiation events. In contrast, the type II loci are distributed across 12 subfamilies that were established before the diversification of the angiosperms. Overall, expressed sequence tag (EST) data exist for 20 of these loci; further characterization of gene expression patterns will be an important next step. This characterization of Aquilegia MADS-box transcription factors thereby lays the foundation for many crucial studies on the development and evolution of Aquilegia as well as the conservation of function across the MADS-box gene family.

Key words: Aquilegia, gene duplication, MADS-box genes, Ranunculaceae.

well known for regulating growth and developmental processes across eukaryotes, but it appears to be especially critical in plants (Messenguy & Dubois, 2003; Gramzow & Theissen, 2010). Members of the family are defined by the presence of the conserved MADS-box, which is typically located at or close to the 5' end of the coding region and consists of a 180 bp motif. This sequence encodes a DNA-binding domain that recognizes regulatory elements known as CArG boxes, which have the consensus sequence 5'CC[A/T]₆GG-3' (Riechmann et al., 1996b). Although MADS-box genes are found in animals, fungi, and plants, they tend to be much more diverse in plants, particularly seed plants (Nam et al., 2003; Gramzow & Theissen, 2010). Broadly speaking, there are two main evolutionary lineages of MADS-box genes, which are referred to as type I and type II (Fig. 1). The better-studied type II lineage includes MEF2like genes in animals and fungi and MIKC-type genes in plants (Alvarez-Buylla et al., 2000). The MIKCtype genes derive their name from the four conserved domains defined in their protein sequences: MADS (M), Intervening (I), Keratin-like (K), and C-terminal (C) (Ma et al., 1991). MIKC-type genes can be further subdivided into the MIKC^c and MIKC* (or Mδ) types, of which the MIKC^c-type are the best-characterized

The MADS-box family of transcription factors is group of MADS-box genes (Gramzow & Theissen, 2010). Numerous studies have demonstrated that MIKC^c MADS-box genes function as dimers and in higher-order protein complexes (reviewed in Gramzow & Theissen, 2010). These protein-protein interactions are primarily mediated by \alpha-helical regions of the K domain with some contributions from the I and MADS domains (Riechmann et al., 1996a; Yang et al., 2003; Yang & Jack, 2004). In contrast to the well-understood M, I, and K domains, the C-terminal domain shows much lower levels of sequence conservation overall and remains somewhat of a mystery. Several subfamilies of MIKC^c loci contain transcriptional activation domains at their Cterminus (Honma & Goto, 2001), but no functions have been clearly ascribed to the highly conserved Cterminal motifs that define each lineage of the MIKC^c subfamily (reviewed in Litt & Kramer, 2010). Of the 14 major angiosperm lineages of MIKC^c loci, 11 contribute directly to the transition to flowering or the development of flowers themselves (reviewed in Yant et al., 2009; Gramzow & Theissen, 2010; Melzer et al., 2010), making comparative studies of this group of particular importance for understanding the evolution of flowering plants. The MIKC*-type were originally found in mosses and clubmosses but have now also been identified in well-studied seed plants

¹ We thank the Kramer lab for critical comments on the manuscript. This work was supported by National Science Foundation (NSF) awards IOS-0720240 and EF-0412727 to E. M. K.

² Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, U.S.A.

^{*}Author for correspondence: ekramer@oeb.harvard.edu. doi: 10.3417/2010131

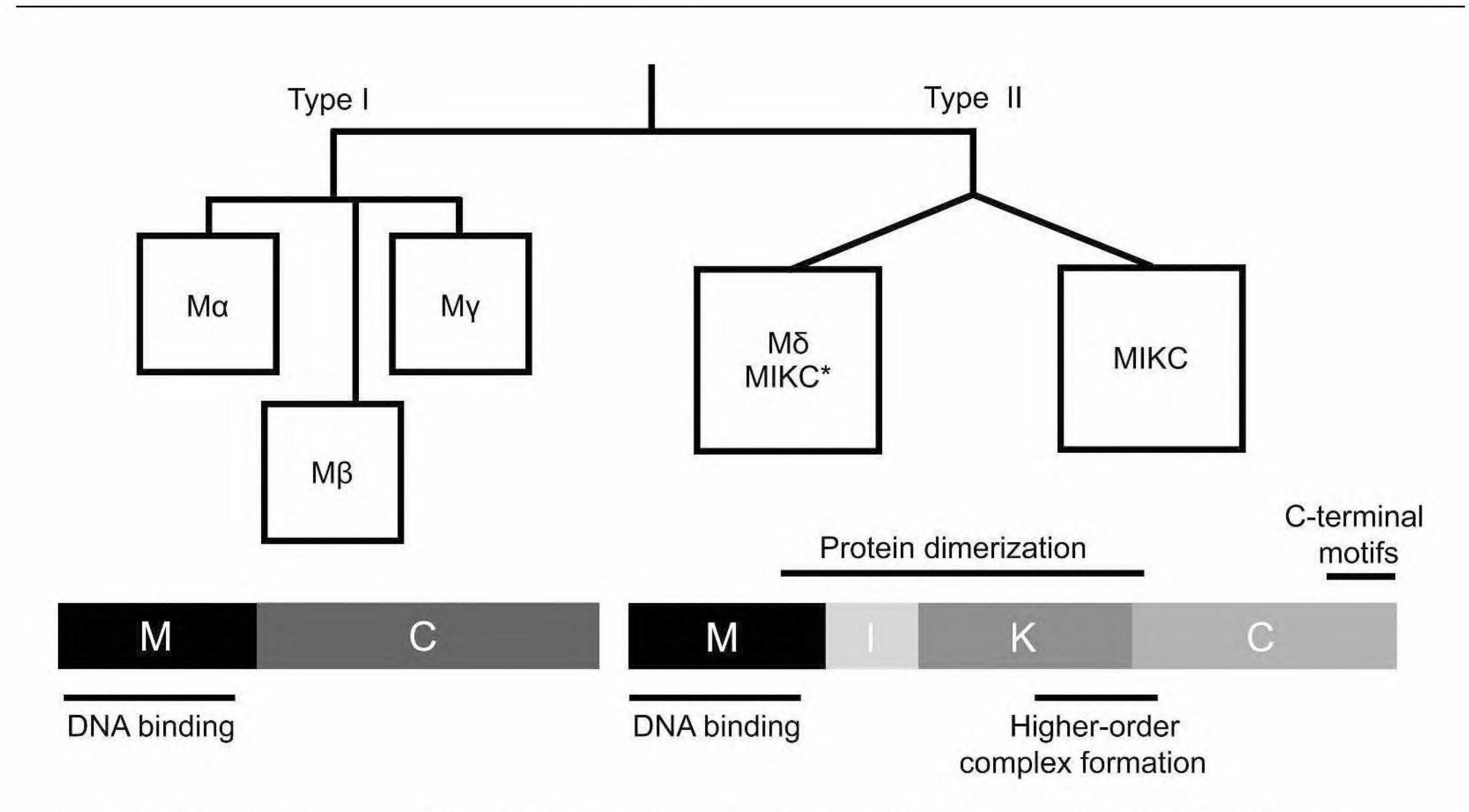


Figure 1. Schematic summary of the evolutionary relationships between the five major subfamilies of MADS-box containing loci. While all of the loci are defined by the presence of the MADS (M) domain, only the type II genes show conservation of three additional domains: Intervening (I), Keratin-like (K), and C-terminal (C). Note that the C-terminal region of the type I loci is completely distinct from that of the type II.

such as the core eudicot Arabidopsis (DC.) Heynh. and grass Oryza L. (reviewed in Zobell et al., 2010). Recently, MIKC* loci have been implicated in microgametophyte maturation and development, but much more work is required to understand whether this is a common feature of the subfamily (Adamczyk & Fernandez, 2009).

Although type I MADS-box genes outnumber type II in the Arabidopsis genome (Parenicova et al., 2003), their evolution and functions are comparatively poorly understood. The type I genes can be subdivided into the Mα, Mβ, and Mγ subfamilies, with Mβ being sister to Mγ (Fig. 1; Parenicova et al., 2003). They are much more diverse in their structures than the type II and lack a canonical K domain, although they do appear to form protein dimers (de Folter et al., 2005; Bemer et al., 2008). Two clear features have emerged regarding the type I MADS-box genes. First, they have experienced a basal eudicot model Aquilegia L. (columbine) in much more rapid birth-and-death evolution than type II homologs (Nam et al., 2004). While phylogenetic analyses of type II loci result in many deeply conserved lineages, type I loci tend to cluster together by taxon, reflecting independent and relatively recent gene duplications (Nam et al., 2004; Arora et al., 2007). Second, they are commonly involved in the development of the female gametophyte and endosperm, as confirmed by both forward genetics and broad expression studies (Bemer et al., 2010b and references therein). These two features are possibly interrelated and may reflect the often rapid evolution

of loci involved in fertilization processes (reviewed in Tian et al., 2009).

Since 2000, the development of high-throughput sequencing has facilitated transcriptome and genome analysis of a wide array of plant species, including Arabidopsis, Oryza, and Vitis L. (Joint Genome Institute, 2010). This work has facilitated evolutionary studies of gene lineage evolution across the angiosperms as well as comparative analysis of functional evolution within this context (e.g., Arora et al., 2007; Bowman et al., 2007). These studies have highlighted the critical interplay between gene duplication and functional divergence, even when primary sequence is highly conserved (e.g., Causier et al., 2005, 2010). To date, this work has primarily focused on the grass and core eudicot model systems, but new sequencing efforts now allow us to add a third major lineage of angiosperms in the form of the Ranunculaceae. Aquilegia consists of ca. 70 perennial species distributed across temperate North America, Europe, and Asia (reviewed in Hodges & Arnold, 1994; Kramer, 2009). These recently diversified species have long fascinated researchers working in the fields of evolution and ecology due to their association of poor genetic differentiation with highly divergent pollinator syndromes (Hodges et al., 2004; Hodges & Derieg, 2009). More recently, Aquilegia has become a model for the evolution of floral morphology thanks to its novel floral organ types, which include first whorl petaloid sepals,

spurred petals in the second whorl, and a unique fifth organ type of sterile staminodia positioned between the fertile stamens and carpels (Kramer et al., 2007; Kramer, 2009). For all these reasons, as well as its relatively small genome size (~300 million basepairs [Mbp] 2C), Aquilegia is currently the subject of extensive genetic and genomic research that has produced an extensive expressed sequence tag (EST) dataset, a physical map, functional tools and, most recently, an 8X genome sequence produced by the Department of Energy (DOE) Joint Genome Institute (Kramer, 2009; Joint Genome Institute, 2010; Kramer & Hodges, 2010). This genome sequence comprises ca. 302 Mbp arranged in 971 scaffolds, of which ~2.9% is gap. In order to get a better understanding of the evolution of MADS-box genes and to create a resource for researchers interested in working with Aquilegia, we used the publically available first assembly of the hybrid A. coerulea E. James 'Origami' genome to identify MADS-box containing loci. The obtained sequences were used in phylogenetic analyses of the entire MADS-box family in order to assign subfamily affinities and were further included in more detailed studies of the MIKC^c subfamily to confirm lineage homology. These findings are discussed in the context of similar studies with particular consideration for the implications of deep patterns of MADS-box gene evolution.

Materials and Methods

IDENTIFICATION OF MADS-BOX GENES FROM THE HYBRID AQUILEGIA COERULEA 'ORIGAMI' GENOME

In order to expand the set of 16 published Aquilegia MADS-box genes (Kramer et al., 2003, 2004, 2007), we used Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1997) to perform a search of the recently released hybrid A. coerulea 'Origami' genome, annotation v1.0 (Joint Genome Institute, 2010) using previously identified MADS domain sequences from Arabidopsis thaliana (L.) Heynh., Vitis vinifera L., and Oryza sativa L. (Parenicova et al., 2003; Arora et al., 2007; Diaz-Riquelme et al., 2009). Specifically, we used the Arabidopsis sequences for AGL16, AGL29, AGL33, AGL39, AGL48, AGL50, AGL58, AGL61, AGL80, AGL82, AGL86, AGL87, AGL97, AGL98, AGL100, AGL101, AGL103, and SEEDSTICK; Vitis sequences for VvAGL12, VvAGL17.1, VvFLC2, and VvTM8; and Oryza sequences for OsMADS62, OsMADS68, Os-MADS89, OsMADS90, OsMADS94, and OsMADS96. Each identified putative Aquilegia coerulea 'Origami' locus was examined for open reading frames using SoftBerry FGENESH (Salamov & Solovyev, 2000). Predicted protein and complementary DNA (cDNA) sequences were extracted and BLASTed back to both GenBank, in order to identify the MADS-box region and make initial assessments of affinity, and to the *Aquilegia* genome sequence itself to identify any other closely related paralogs. New MADS-box gene sequences were deposited in GenBank under accession numbers JX680222–JX680256 (Table 1).

Confirmed MADS-box containing loci were phylo-

PHYLOGENETIC ANALYSES

genetically analyzed in order to determine their membership in the type I versus type II subfamilies. This required using ClustalW (Larkin et al., 2007) to construct an amino acid sequence alignment of the ~60 residue MADS domain. In addition to all of the Aquilegia sequences, this alignment included all Arabidopsis and Oryza MADS loci as well as Vitis VvTM8 and Solanum lycopersicum L. TM8 to represent the TM8 lineage (see Arora et al., 2007; Diaz-Riquelme et al., 2009 for all accession numbers). Neighbor-joining (NJ) analysis was used on this dataset (Saitou & Nei, 1987) as implemented by PAUP* (Swofford, 2002). The NJ phylogeny (Fig. 2) was rooted along the branch separating the type II sequences (MIKC^c and MIKC*/Mδ) from those of type I, in keeping with previous studies (Alvarez-Buylla et al., 2000; Nam et al., 2004). The MIKC^c loci were further analyzed in order to determine specific affinities with deeply conserved lineages. This analysis used an amino acid alignment encompassing the M, I, and K domains (collectively termed MIK) of Aquilegia, Arabidopsis, Oryza, Petunia Juss., and Vitis MIKC^c representatives (for accession numbers see Table 1; Immink et al., 2003; Parenicova et al., 2003; Arora et al., 2007; Diaz-Riquelme et al., 2009). The completed alignment included 130 loci and 175 residues (contact author for alignments). Maximum likelihood (ML) phylogenetic analyses were performed using RAxML (Stamatakis et al., 2005; Stamatakis, 2014) as implemented by the CIPRES portal (Miller et al., 2009). For the purposes of the RAxML analyses, the best protein model of evolution was JTT (Jones) based on MrBayes 3.1 (Ronquist & Huelsenbeck, 2003; Huelsenbeck et al., 2008) amino acid mixed model tests (greater than 99 posterior probability [PP]). Branch support was estimated by performing 1000 replicates of fast bootstrapping (Stamatakis et al., 2008) using the same parameters as the original analysis. The TM8 lineage was used to root this phylogeny, based both on the results of the analysis of the MADS domain alone and those of previous studies (Becker & Theissen, 2003). Matrixes and

Table 1. Hybrid Aquilegia coerulea E. James 'Origami' MADS-box genes with genome location. Expressed sequence tag (EST) numbers are for the Dana Farber Cancer Institute A. formosa Fisch. ex DC. \times A. pubescens Coville database.

Locus	Location					
	Scaffold	MADS Pos	Strand	Type	EST	GenBank
AqAGL60	3	9581086	Plus	Alpha	none	JX680221
AqAGL61	82	280863	Minus	Alpha	none	JX680222
AqAGL62	31	2417989	Plus	Alpha	none	JX680223
AqAGL63	24	3291956	Minus	Alpha	none	JX680224
AqAGL64	96	478666	Plus	Alpha	none	JX680225
AqAGL65	2	5042310	Plus	Delta	TC33318	JX680226
AqAGL66	18	2220081	Minus	Alpha	none	JX680227
AqAGL67	18	3429116	Plus	Alpha	none	JX680228
AqAGL68	69	404345	Minus	Alpha	none	JX680229
AqAGL69	69	400144	Minus	Alpha	none	JX680230
AqAGL70	15	1987421	Plus	Alpha	none	JX680231
AqAGL71	52	729783	Plus	Alpha	none	JX680232
AqAGL72	41	479166	Plus	Alpha	none	JX680233
AqAGL73	7	190653	Minus	Alpha	none	JX680234
AqAGL74	5	5460796	Plus	Alpha	none	JX680235
AqAGL75	54	1514596	Plus	Alpha	none	JX680236
AqAGL80		4403062	Minus	Gamma	none	JX680237
AqAGL81	1529	2159	Minus	Gamma	none	JX680238
AqAGL82	112	22986	Plus	Gamma	none	JX680239
AqAGL83	29	109423	Minus	Gamma	none	JX680240
AqAGL84	7	5134489	Plus	Gamma	none	JX680241
AqAGL85	27	969516	Plus	Gamma	none	JX680242
AqAGL86	18	1770226	Minus	Beta/gamma	none	JX680243
AqAP3-I	6	2204058	Plus	MIKC	TC22599	EF489478
AqAP3-2	6	2147564	Plus	MIKC	TC24405	EF489477
AqAP3-3	7	3032344	Plus	MIKC	TC20315	EF489476
AqAP3-3b	38	1797328	Plus	MIKC	none*	H0694798
AqPI	14	1712364	Minus	MIKC	TC21654	EF489475
AqBS	22	2313113	Minus	MIKC	none*	AY436713
AqSEPI	6	8363818	Minus	MIKC	TC30455	JX680244
AqSEP2A	2	6945234	Minus	MIKC	TC23935	JX680245
AqSEP2B	2	6911967	Minus	MIKC	none	JX680246
AqSEP3		2306215	Plus	MIKC	TC20920	JX680247
AqAGL6	13	3100284	Minus	MIKC	TC27019	JX680248
AqAGL17	10	436736	Minus	MIKC	none	JX680249
AqAGL12	14	490009	Plus	MIKC	none	JX680251
AqAGL15	9	1871509	Minus	MIKC	TC30235	JX680250
AqAGI	136	79379	Minus	MIKC	TC22246	AY464111
AqAG2	22	731674	Plus	MIKC	none*	AY464110
AqAGL24.1	15	1246754	Minus	MIKC	TC24816	HQ173338
AqAGL24.2	7	4643609	Plus	MIKC	TC33172**	HQ173339
AqFLIA	2	6930012	Minus	MIKC	TC23520	JX680252
AqFLIB	2	6899144	Minus	MIKC	TC27021	JX680253
AqSOC1.1	13	3057695	Plus	MIKC	DR913118	HQ173336
AqSOCI.2	35	2391819	Minus	MIKC	TC31575	JX680254
AqSOC1.3	3	9953490	Plus	MIKC	none	JX680255
AqSOC1.4		8820680	Minus	MIKC	none	JX680256

^{*} Locus is not represented in the EST database, but expression has been confirmed using reverse transcriptase (RT)-polymerase chain reaction (PCR).

** EST is incorrectly spliced.

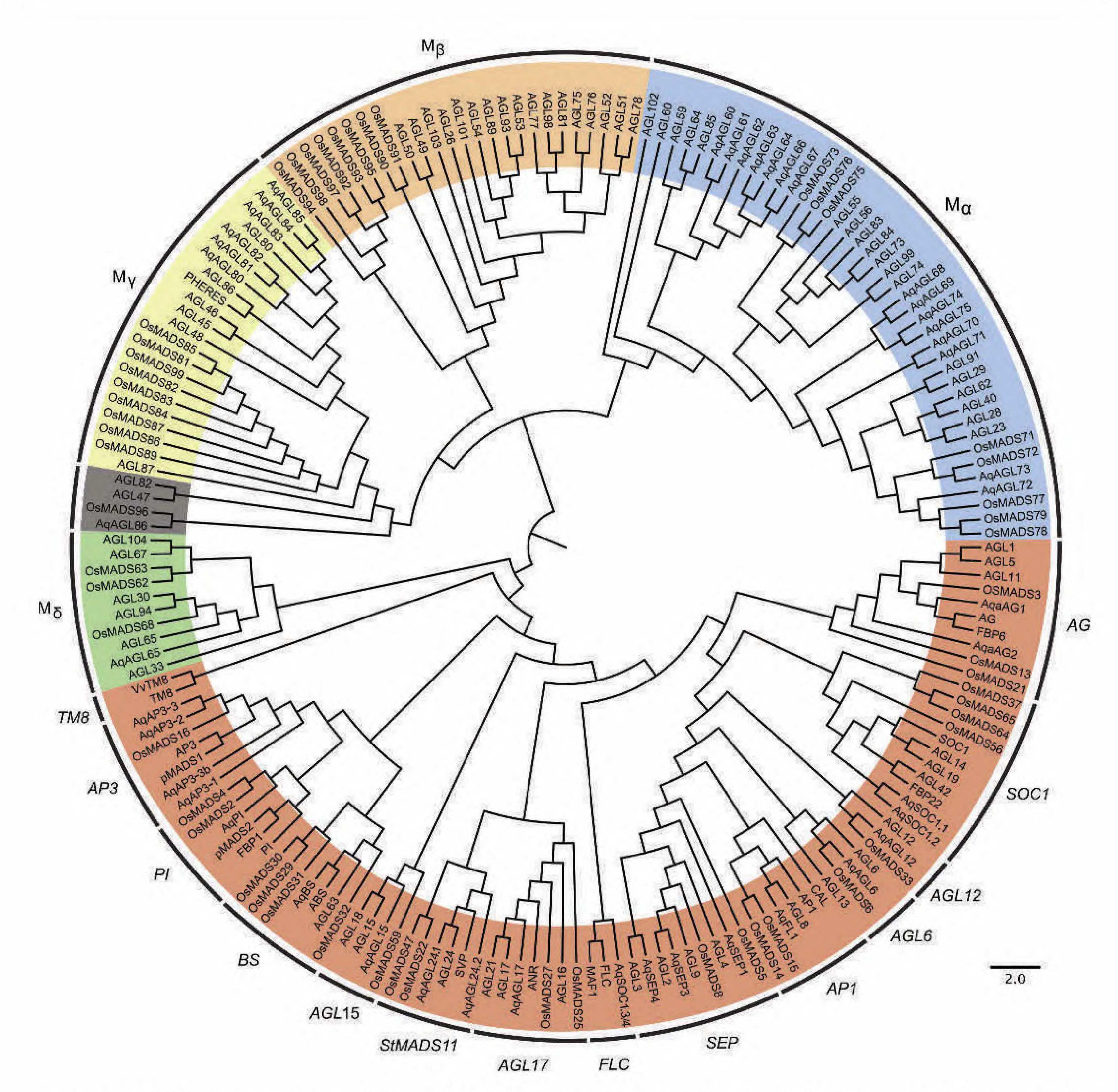


Figure 2. Neighbor-joining (NJ) analysis of MADS domain sequences from all identified Aquilegia, Arabidopsis, and Oryza loci. Specific lineages are indicated by colors and bracketing. The type I lineages are the Mγ (yellow), Mβ (orange), and Mα (blue), while the type II are the MIKC^c (red, with individual lineages denoted by brackets) and MIKC* (or Mδ; green). A paraphyletic group of four loci are colored in gray. In addition to the new AqAGL86 sequence, these include three sequences (AGL47, AGL82, and OsMADS86) that have previously been placed with Mβ but are instead associated with Mγ in our analysis.

trees associated with this study were deposited in study/TB2:S13212>).

Results and Discussion

THE MADS-BOX FAMILY OF THE HYBRID AQUILEGIA COERULEA 'ORIGAMI'

We have identified 47 MADS domain containing loci in the recently sequenced genome of the hybrid Aquilegia coerulea 'Origami' (Table 1; Fig. 2). These genes are distributed across 29 different scaffolds that range in size from almost 2 Mbps to 355 kilo basepaird (Kbps). Scaffolds 2, 3, 4, 6, 7, 13, 14, 15,

18, 22, and 69 contain multiple MADS-box loci, but TreeBase (http://purl.org/phylo/treebase/phylows/ only scaffolds 2, 18, and 69 appear to represent tandem duplications (see below for further discussion). Note that we detected two scaffolds that appear to have assembly errors: 136, which has two identical tandem copies of AqAGI, and 96, which has two identical tandem copies of AqAGL64. These duplicates, which we believe to be artificial, were not included in the analysis.

> An NJ analysis of all of the recovered MADS domains demonstrates that 23 loci distributed across 18 scaffolds fall into type I, while 24 loci distributed across 15 scaffolds are placed in type II. Overall, the phylogenetic tree typology of MADS domain se-

quences from all identified Arabidopsis, Aquilegia, and Oryza loci (Fig. 2) is largely consistent with previous studies (Parenicova et al., 2003; Arora et al., 2007). The one inconsistency is the placement of AGL47, AGL82, and OsMADS96 relative to the MB and My subfamilies, which will be further discussed below. The total number of MADS domain loci identified, 47, is considerably less than the 107 and 75 known from Arabidopsis and Oryza, respectively (Parenicova et al., 2003; Arora et al., 2007). This is likely due to two factors. First, both Arabidopsis and Oryza have experienced genome duplication events in their relatively recent genomic history (De Bodt et the question of whether the Mß lineage is truly al., 2005), which may have increased the numbers of MADS-box loci. Second, this study was conducted using the v1.0 annotation of the hybrid Aquilegia coerulea 'Origami' genome, meaning that further annotation may yet identify additional loci.

TYPE I MADS-BOX GENES

The Aquilegia type I clade contains three monophyletic lineages roughly corresponding to the previously defined Mα, Mβ, and Mγ (Fig. 2). The Mα clade includes 15 Aquilegia representatives, which appear to define at least three separate lineages that are largely independently diversified relative to the Arabidopsis and Oryza representatives. The My clade contains six Aquilegia representatives that are, again, likely to be independently radiated from the other identified loci. The one point of disagreement between our analysis and previous studies is the placement of the Arabidopsis sequences AGLA7 and AGL82 and the Oryza OsMADS96. Parenicova et al. (2003) placed AGL47 and AGL82 in the Mβ clade, albeit with no support. Likewise, the analysis of Arora et al. (2007) identified OsMADS96 as an MB representative but with no reported support. In our analysis, these three loci fall out with a new Aquilegia sequence, AqAGL86, as paraphyletic to the My clade rather than with the M\beta. Closer inspection of the four complete sequences reveals no obvious shared motifs, either among AGL47/82, OsMADS96, and AqAGL86 or between these genes and either the MB or My homologs (Parenicova et al., 2003; Arora et al., 2007; and data not shown). Given that our analysis similarly lacks support for these relationships, we cannot make strong conclusions beyond saying that AqAGL86 is currently associated with the My clade.

This raises the larger question, however, of how conserved the MB lineage really is across angiosperms. Previous studies held that MB representatives were specific to the Brassicaceae (Leseberg et al., 2006), but Arora et al. (2007) recovered an apparent clade of Oryza Mß loci. Likewise, in our NJ

analysis, most of these same Oryza genes are associated with the original MB genes from Arabidopsis (Fig. 2). It is interesting to note, however, that no clear MB representatives have been recovered from the Aquilegia genome yet. Furthermore, examination of the putative MB Arabidopsis and Oryza loci does not reveal any obvious shared motifs, either within the MADS domain or outside it (Parenicova et al., 2003; Arora et al., 2007). Given this rather weak association between the Arabidopsis and Oryza MB loci, along with their apparent absence from the Aquilegia genome, it may be necessary to re-examine conserved across the angiosperms. Of course, the annotation of the Aquilegia genome is in its early stages, and Mß loci may yet be discovered.

In terms of patterns of genomic structure, most of the type I genes are relatively dispersed across different scaffolds. There are three pairs—AqAGL66/ 67, AqAGL68/69, and AqAGL84/85—that represent interesting cases. Each one of these pairs has identical or almost identical MADS domain sequences, while the rest of the coding regions contain a small number of clear differences. Thus, we have annotated them as separate loci, but it is likely that they are derived from relatively recent duplication events. One of the pairs, AqAGL68/69, is in fact close together on the same scaffold (69), suggesting a recent tandem duplication. However, the other two pairs are not close together, with AqAGL66/67 on the same scaffold (18) but 1.2 Mbp apart and AqAGL84/ 85 on completely different scaffolds (7 and 27). Although it is possible that scaffolds 7 and 27 will ultimately be joined into one chromosomal unit, these scaffolds are approximately 6.1 and 3 Mbp, respectively, so the loci are at least 2 Mbp apart based on their locations in the scaffolds. Neither of these pairs shows evidence of shared synteny that would suggest a large-scale duplication event. As is typical for type I loci (De Bodt et al., 2003), the Aquilegia representatives are predicted to contain few if any introns, with only four loci predicted to have either one (AqAGL60, AqAGL63, AqAGL69) or two (AqAGL73) introns. It is interesting to note in this regard that for the apparent tandem duplication pair AqAGL68/69, the former lacks introns while the latter has one, possibly reflecting a retroduplication origin for AqAGL68.

TYPE II MADS-BOX GENES

Of the 24 type II MADS-box genes, only one member is in the MIKC* subfamily, AqAGL65, a homolog of the P-clade (Nam et al., 2004), with the balance in the MIKC^c. Many of the Aquilegia MIKC^c

members have been previously described, particularly in regard to their potential roles in novel floral organ identity in Aquilegia (Kramer et al., 2003, 2004, 2007; Sharma et al., 2011), but this is the first report for six of the loci (AqAGL12, AqAGL15, AqAGL17, AqSOC1.2, AqSOC1.3, AqSOC1.4). Of these new loci, AqAGL15 and AqSOC1.2 are also represented by ESTs in the A. formosa Fisch. ex DC. X A. pubescens Coville databases, but expression of the remaining loci has not yet been demonstrated (Table 1). There is one previously published locus, AqFL2, which was originally isolated from A. vulgaris L., but we have been unable to identify it in the hybrid A. coerulea 'Origami' genome. A partial coding region for AqFL2 was defined based on four identical cDNA fragments that were obtained in the process of cloning the full-length AqFL1 cDNA. Given that AqFL2 appears to be a representative of an ancient paralogous FUL-like lineage in the Ranunculales (Litt & Irish, 2003), we are inclined to believe that it was not a spurious identification, but the possibility exists that it has either been lost from the hybrid A. coerulea 'Origami' genome or has not been covered by current sequencing. AqSOC1.3/AqSOC1.4 are also of particular note since they lack introns and, thus, appear to be retroduplications. The two open reading frames are almost identical but are located on different scaffolds (3 and 5, respectively) with different neighboring loci, including predicted transposon sequence flanking AqSOC1.4, which further supports the retroduplication hypothesis. Although AqSOC1.3/AqSOC1.4 are associated with FLC in the MADS domain analysis (Fig. 2), this is not supported by the MIK analysis (see below). Another interesting point is that AqFL1 and AqSEP2 are each represented twice in the genome, being part of a large segmental duplication on scaffold 2. We term these loci AqFL1A/B and AqSEP2A/B. Both homeologous pairs contain introns of different lengths but have only a very small number of differences in their coding regions. It appears that ESTs from three of the four loci are present in the A. formosa \times A. pubescens database (Table 1), with AqSEP2B remaining to be confirmed as an expressed locus.

In order to better understand the phylogenetic relationships among *Aquilegia* MIKC^c loci, we created an amino acid alignment covering the MIK domains, which can be confidently aligned across the entire subfamily. We expanded sampling in this dataset to include *Petunia* and *Vitis* homologs and analyzed it using ML as implemented by RAxML software (Fig. 3). The topology of the resultant phylogeny is largely consistent with previous studies (Becker & Theissen, 2003; Gramzow & Theissen,

2010). Although AqSOC1.3/AqSOC1.4 are associated with the FLC lineage (Fig. 2), they fall into the SOCI clade with strong support in the MIK analysis (Fig. 3A). This reflects the fact that while AqSOC1.3/ AqSOC1.4 have rather divergent MADS domains, their I-, K-, and C-terminal domains contain synapomorphic motifs for the SOCI subfamily. Unlike AqSOC1.3/AqSOC1.4, the other subfamily members, AqSOC1.1/AqSOC1.2 have the typical six introns associated with MIKC^c loci. Therefore, no Aquilegia representatives have been identified for the FLC or TM8 lineages, highlighting the mysterious nature of both. FLC is notable because although it is a highly pleiotropic locus in Arabidopsis, affecting vernalization response, temperature-dependent germination, water use, and phase change (McKay et al., 2003; Alexandre & Hennig, 2008; Chiang et al., 2009; Willmann & Poethig, 2011), orthologs have yet to be identified outside the core eudicots (Becker & Theissen, 2003; Gramzow & Theissen, 2010). Despite some possible evidence for a conserved role in flowering time response (Reeves et al., 2007), no direct functional data exist for FLC orthologs in other core eudicots, and the source of their derivation remains unclear. One possibility is that the lineage was derived from the y hexaploidization event at the base of the core eudicots (Jiao et al., 2012; Vekemans et al., 2012), but even if that is the case, it remains to be determined what the most closely related lineages might be and which aspects of the complex functional repertoire in Arabidopsis might be conserved across core eudicots. The TM8 lineage is even more enigmatic. Very few homologs have been identified, the majority of which are found in the core eudicots (although Arabidopsis lacks a TM8 ortholog; Becker & Theissen, 2003), and no function has yet been ascribed to any member. The ongoing, extensive transcriptomic and genomic studies of diverse angiosperms will hopefully help answer some of these questions.

CONCLUSION

Aquilegia is an important new model system for the study of both ancient and recent evolutionary processes. Our identification of a large number of MADS-box containing loci will aid comparative studies seeking to bridge the gap between grass and core eudicot models. In particular, the characterization of a large number of type I MADS-box genes will allow researchers to determine whether the novel expression patterns and functions associated with these loci are deeply conserved across the angiosperms. Overall, our finding that the Aquilegia type I and type II subfamilies have very different evolu-

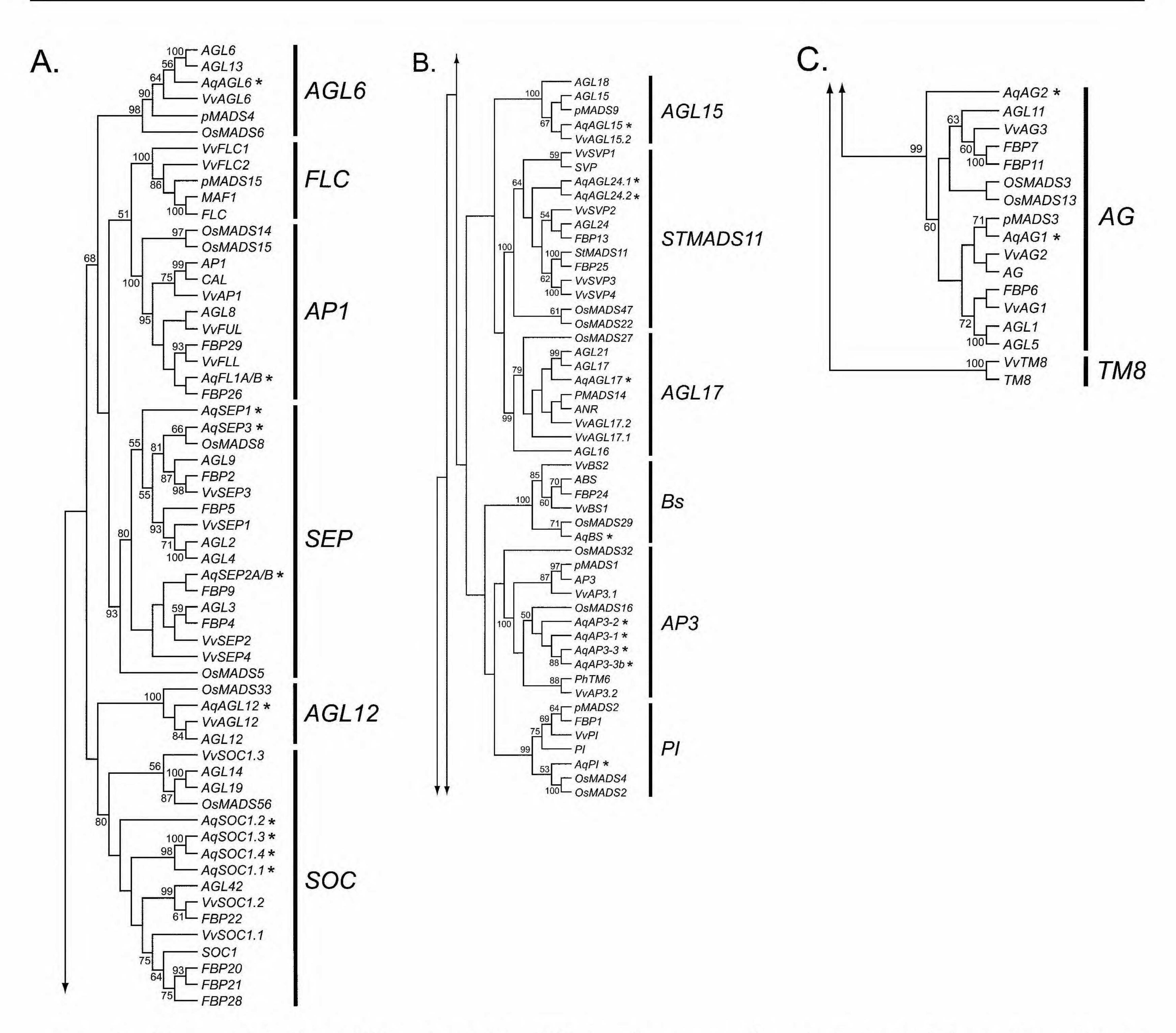


Figure 3. Maximum Likelihood (ML) analysis of the MIK domain sequences from *Aquilegia*, *Arabidopsis*, *Oryza*, *Petunia*, and *Vitis*, and MIKC^c loci. Bootstrap support values of more than 50% are indicated at nodes. Brackets on the right denote specific lineage affiliations. Asterisks indicate *Aquilegia* loci.

tionary histories is consistent with similar studies in *Arabidopsis*, *Petunia*, and *Oryza* (Immink et al., 2003; Parenicova et al., 2003; Arora et al., 2007). On the other hand, although there are some examples of seemingly recent gene duplication events across these loci, there appear to be fewer tandem paralogs, especially among the type I loci, than has been observed in other models (De Bodt et al., 2003; Nam et al., 2004; Bemer et al., 2010a). This seems to be consistent with an overall smaller number of MADS-box genes in *Aquilegia*.

Literature Cited

Adamczyk, B. J. & D. E. Fernandez. 2009. MIKC* MADS domain heterodimers are required for pollen maturation and tube growth in *Arabidopsis*. Pl. Physiol. (Lancaster) 149: 1713–1723.

Alexandre, C. M. & L. Hennig. 2008. FLC or not FLC: The other side of vernalization. J. Exp. Bot. 59: 1127–1135.

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. H. Zhang, Z. Zhang, W. Miller & D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucl. Acids Res. 25: 3389–3402.

Alvarez-Buylla, E. R., S. Pelaz, S. J. Liljegren, S. E. Gold, C. Burgeff, G. S. Ditta, L. Ribas de Pouplana, L. Martinez-Castilla & M. F. Yanofsky. 2000. An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. Proc. Natl. Acad. Sci. U.S.A. 97: 5328–5333.

Arora, R., P. Agarwal, S. Ray, A. K. Singh, V. P. Singh, A. K. Tyagi & S. Kapoor. 2007. MADS-box gene family in rice: Genome-wide identification, organization and expression profiling during reproductive development and stress. BMC Genomics 8: 242.

Becker, A. & G. Theissen. 2003. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. Molec. Phylogen. Evol. 29: 464–489.

Bemer, M., M. Wolters-Arts, U. Grossniklaus & G. C. Angenent. 2008. The MADS domain protein DIANA acts

- together with AGAMOUS-LIKE80 to specify the central cell in Arabidopsis ovules. Pl. Cell 20: 2088–2101.
- Bemer, M., J. Gordon, K. Weterings & G. C. Angenent. 2010a. Divergence of recently duplicated M gamma-type MADS-box genes in Petunia. Molec. Biol. Evol. 27: 481-495.
- Bemer, M., K. Heijmans, C. Airoldi, B. Davies & G. C. Angenent. 2010b. An atlas of type I MADS-box gene expression during female gametophyte and seed development in Arabidopsis. Pl. Physiol. (Lancaster) 154: 287-300.
- Bowman, J. L., S. K. Floyd & K. Sakakibara. 2007. Green genes—Comparative genomics of the green branch of life. Cell (Cambridge) 129: 229–234.
- Causier, B., R. Castillo, J. L. Zhou, R. Ingram, Y. B. Xue, Z. Schwarz-Sommer & B. Davies. 2005. Evolution in action: Following function in duplicated floral homeotic genes. Curr. Biol. 15: 1508–1512.
- Causier, B., R. Castillo, Y. B. Xue, Z. Schwarz-Sommer & B. Davies. 2010. Tracing the evolution of the floral homeotic B- and C-function genes through genome synteny. Molec. Biol. Evol. 27: 2651–2664.
- Chiang, G. C. K., D. Barua, E. M. Kramer, R. M. Amasino & K. Donohue. 2009. Major flowering time gene, FLOWERING LOCUS C, regulates seed germination in Arabidopsis thaliana. Proc. Natl. Acad. Sci. U.S.A. 106: 11,661–11,666.
- De Bodt, S., J. Raes, K. Florquin, S. Rombauts, P. Rouzé, G. Theissen & Y. Van de Peer. 2003. Genomewide structural annotation and evolutionary analysis of the type I MADS-box genes in plants. J. Molec. Evol. 56: 573-586.
- De Bodt, S., S. Maere & Y. Van de Peer. 2005. Genome duplication and the origin of angiosperms. Trends Ecol. Evol. 20: 591–597.
- de Folter, S., R. G. H. Immink, M. Kieffer, L. Parenicova, S. R. Henz, D. Weigel, M. Busscher, M. Kooiker, L. Colombo, M. M. Kater, B. Davies & G. C. Angenent. 2005. Comprehensive interaction map of the Arabidopsis MADS-box transcription factors. Pl. Cell 17: 1424–1433.
- Diaz-Riquelme, J., D. Lijavetzky, J. M. Martinez-Zapater & M. J. Carmona. 2009. Genome-wide analysis of MIKC^ctype MADS-box genes in grapevine. Pl. Physiol. (Lancaster) 149: 354–369.
- Gramzow, L. & G. Theissen. 2010. A hitchhiker's guide to the MADS world of plants. Genome Biol. 11: 214.
- Hodges, S. A. & M. L. Arnold. 1994. Columbines—A geographically widespread species flock. Proc. Natl. Acad. Sci. U.S.A. 91: 5129-5132.
- Hodges, S. A. & N. J. Derieg. 2009. Adaptive radiations: McKay, J. K., J. H. Richards & T. Mitchell-Olds. 2003. From field to genomic studies. Proc. Natl. Acad. Sci. U.S.A. 106: 9947–9954.
- Hodges, S. A., M. Fulton, J. Y. Yang & J. B. Whittall. 2004. Verne Grant and evolutionary studies of Aquilegia. New Phytol. 161: 113-120.
- Honma, T. & K. Goto. 2001. Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. Nature 409: 525-529.
- Huelsenbeck, J. P., P. Joyce, C. Lakner & F. Ronquist. 2008. Bayesian analysis of amino acid substitution models. Philos. Trans., Ser. B 363: 3941–3953.
- Immink, R. G. H., S. Ferrario, J. Busscher-Lange, M. Kooiker, M. Busscher & G. C. Angenent. 2003. Analysis of the petunia MADS-box transcription factor family. Molec. Genet. Genomics 268: 598-606.

- Jiao, Y., J. Leebens-Mack, S. Ayyampalayam, J. E. Bowers, M. R. McKain, J. McNeal, M. Rolf, D. R. Ruzicka, E. Wafula, N. J. Wickett, X. Wu, Y. Zhang, J. Wang, Y. Zhang, E. J. Carpenter, M. K. Deyholos, T. M. Kutchan, A. S. Chanderbali, P. S. Soltis, D. W. Stevenson, R. McCombie, J. C. Pires, G. K.-S. Wong, D. E. Soltis & C. W. dePamphilis. 2012. A genome triplication associated with early diversification of the core eudicots. Genome Biol. 13: R3.
- Joint Genome Institute. 2010. Phytozome, ver. 6.0. http:// www.phytozome.net/>, accessed 10 December 2010.
- Kramer, E. M. 2009. Aquilegia: A new model for plant development, ecology, and evolution. Annual Rev. Pl. Biol. 60: 261–277.
- Kramer, E. M. & S. A. Hodges. 2010. Aquilegia as a model system for the evolution and ecology of petals. Philos. Trans., Ser. B 365: 477–490.
- Kramer, E. M., V. S. Di Stilio & P. Schluter. 2003. Complex patterns of gene duplication in the APETALA3 and PISTILLATA lineages of the Ranunculaceae. Int. J. Pl. Sci. 164: 1–11.
- Kramer, E. M., M. A. Jaramillo & V. S. Di Stilio. 2004. Patterns of gene duplication and functional evolution during the diversification of the AGAMOUS subfamily of MADS-box genes in angiosperms. Genetics 166: 1011-1023.
- Kramer, E. M., L. Holappa, B. Gould, M. A. Jaramillo, D. Setnikov & P. Santiago. 2007. Elaboration of B gene function to include the identity of novel floral organs in the lower eudicot Aquilegia. Pl. Cell 19: 750–766.
- Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson & D. G. Higgins. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23: 2847–2948.
- Leseberg, C. H., A. Li, H. Kang, M. Duvall & L. Mao. 2006. Genome-wide analysis of the MADS-box gene family in Populus trichocarpa. Gene 378: 84–94.
- Litt, A. & V. F. Irish. 2003. Duplication and diversification in the APETALA1/FRUITFULL floral homeotic gene lineage: Implications for the evolution of floral development. Genetics 165: 821-833.
- Litt, A. & E. M. Kramer. 2010. The ABC model and the diversification of floral organ identity. Seminars Cell Developm. Biol. 21: 129-137.
- Ma, H., M. F. Yanofsky & E. M. Meyerowitz. 1991. AGL1-AGL6, an Arabidopsis gene family with similarity to floral homeotic and transcription factor genes. Genes Developm. 5: 484–495.
- Genetics of drought adaptation in Arabidopsis thaliana: I. Pleiotropy contributes to genetic correlations among ecological traits. Molec. Ecol. 12: 1137–1151.
- Melzer, R., Y. Q. Wang & G. Theissen. 2010. The naked and the dead: The ABCs of gymnosperm reproduction and the origin of the angiosperm flower. Seminars Cell Developm. Biol. 21: 118-128.
- Messenguy, F. & E. Dubois. 2003. Role of MADS-box proteins and their cofactors in combinatorial control of gene expression and cell development. Gene 316: 1–21.
- Miller, M. A., M. T. Holder, R. Vos, P. E. Midford, T. Liebowitz, L. Chan, P. Hoover & T. Warnow. 2009. The CIPRES Science Gateway. CIPRES, ver. 3.1. http:// www.phylo.org/sub_sections/portal>, accessed 14 February 2014.

- Nam, J., C. W. dePamphilis, H. Ma & M. Nei. 2003. Antiquity and evolution of the MADS-box gene family controlling flower development in plants. Molec. Biol. Evol. 20: 1435–1447.
- Nam, J., J. Kim, S. Lee, G. H. An, H. Ma & M. S. Nei. 2004. Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms. Proc. Natl. Acad. Sci. U.S.A. 101: 1910–1915.
- Parenicova, L., S. de Folter, M. Kieffer, D. S. Horner, C. Favalli, J. Busscher, H. E. Cook, R. M. Ingram, M. M. Kater, B. Davies, G. C. Angenent & L. Colombo. 2003. Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: New openings to the MADS world. Pl. Cell 15: 1538–1551.
- Reeves, P. A., Y. H. He, R. J. Schmitz, R. M. Amasino, L. W. Panella & C. M. Richards. 2007. Evolutionary conservation of the *FLOWERING LOCUS C*-mediated vernalization response: Evidence from the sugar beet (*Beta vulgaris*). Genetics 176: 295–307.
- Riechmann, J. L., B. A. Krizek & E. M. Meyerowitz. 1996a. Dimerization specificity of *Arabidopsis* MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. Proc. Natl. Acad. Sci. U.S.A. 93: 4793–4798.
- Riechmann, J. L., M. Wang & E. M. Meyerowitz. 1996b. DNA-binding properties of *Arabidopsis* MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS. Nucl. Acids Res. 24: 3134–3141.
- Ronquist, F. & J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–1574.
- Saitou, N. & M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molec. Biol. Evol. 4: 406–425.
- Salamov, A. A. & V. V. Solovyev. 2000. *Ab initio* gene finding in *Drosophils* genomic DNA. Genome Res. 10: 516–522.
- Sharma, B., C. Guo, H.-Z. Kong & E. M. Kramer. 2011. Petal-specific subfunctionalization of an *APETALA3* paralog in the Ranunculaceae and its implications for petal evolution. New Phytol. 190: 870–883.

- Stamatakis, A. 2014. RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. Bioinformatics. doi: 10.1093/bioinformatics/btu033.
- Stamatakis, A., M. Ott & T. Ludwig. 2005. RAxML-OMP: An efficient program for phylogenetic inference on SMPs. Pp. 288–302 in 8th International Conference on Parallel Computing Technologies (PaCT 2005). Krasnoyarsk, Russia, September 5–9, 2005. Proceedings. Springer-Verlag, Berlin.
- Stamatakis, A., P. Hoover & J. Rougemont. 2008. A rapid bootstrap algorithm for the RAxML web servers. Syst. Biol. 57: 758–771.
- Swofford, D. L. 2002. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), 4.0b10. Sinauer Associates, Sunderland, Massachusetts.
- Tian, X., G. Pascal, S. Fouchécourt, P. Pontarotti & P. Monget. 2009. Gene birth, death, and divergence: The different scenarios of reproduction-related gene evolution. Biol. Reprod. 80: 616–621.
- Vekemans, D., S. Proost, K. Vanneste, H. Coenen, T. Viaene, P. Ruelens, S. Maere, Y. Van de Peer & K. Geuten. 2012. Gamma paleohexaploidy in the stemlineage of core eudicots: Significance for MADS-box gene and species diversification. Molec. Biol. Evol. 29(12): 3793–3806.
- Willmann, M. R. & R. S. Poethig. 2011. The effect of the floral repressor *FLC* on the timing and progression of vegetative phase change in *Arabidopsis*. Development 138: 677–685.
- Yang, Y. & T. Jack. 2004. Defining subdomains of the K domain important for protein-protein interactions of plant MADS proteins. Pl. Molec. Biol. 55: 45–59.
- Yang, Y., L. Fanning & T. Jack. 2003. The K domain mediates heterodimerization of the *Arabidopsis* floral organ identity proteins, APETALA3 and PISTILLATA. Pl. J. 33: 47–59.
- Yant, L., J. Mathieu & M. Schmid. 2009. Just say no: Floral repressors help *Arabidopsis* bide the time. Curr. Opin. Pl. Biol. 12: 580–586.
- Zobell, O., W. Faigl, H. Saedler & T. Munster. 2010. MIKC* MADS-box proteins: Conserved regulators of the gametophytic generation of land plants. Molec. Biol. Evol. 27: 1201–1211.